

Lesson 7

Special Analytic Issues

Instructor's Guide Form

Lesson Title: Special analytic issues

Lesson Goal: For each student to be able to describe the specific issues regarding data obtained in public health surveillance, how to select the proper analytic methods, and how these characteristics affect analyses.

Lesson Objectives: By the end of the lesson, the student will be able to:

- 1) describe the nature of public health surveillance data
- 2) define the nomenclature for variations in health events
- 3) demonstrate correct use of analytic and graphical methods to correct for aberrations in time and space
- 4) demonstrate assessment of completeness of a surveillance system, i.e. validate surveillance information
- 5) select appropriate analytic methods
- 6) describe emergent analytic methods in analysis of surveillance data

Lesson 7

Special Analytic Issues

Instructor's Guide Form (continued)

Equipment and materials needed:

- Overhead projector
- Transparencies #7.1 - #7.13

Time required: 60 minutes

Synopsis of lesson: This lesson describes the characteristics of data obtained in a public health surveillance system affecting analyses.

Adult Education Application: In this chapter, there are a number of concepts that are based on arithmetic principles. For learners who have minimal experience using mathematics, these concepts can be very confusing. If your assessment of the learners is that they will never have to do the calculations, help them grasp the idea by providing a number of practical examples where the concept is helpful in analyzing the public health condition. You could involve them in a discussion about how this concept could be useful to them in their own work environment. For those learners who are more advanced and should be able to do the mathematics, you could provide them with some problems that could be used for practice exercises.

Lesson 7

Special Analytic Issues

Topical Outline

- I. Nature of public health surveillance data**
 - A. Characteristics of surveillance data

- II. Clustering of health events**
 - A. Variations in health events
 - B. Nomenclature for variations

- III. Aberrations in time**
 - A. Graph of current and past experience
 - B. Time-series methods
 - C. Scan statistic

- IV. Aberrations in space and time**
 - A. One method
 - B. Knox's method - an alternative method
 - C. Criticism of Knox's method
 - D. In-class demonstration/exercise

- V. Completeness of coverage**
 - A. Approaches to assessment of completeness
 - B. In-class demonstration/exercise

Lesson 7

Special Analytic Issues

Topical Outline (continued)

VI. Selection of analytic methods

- A. Purpose of surveillance
- B. Purpose of analytic method
- C. Selection of conditions for monitoring
- D. Units of analyses
- E. Provision for updating/correcting data
- F. Determination of baseline
- G. Dealing with outbreaks in baseline
- H. Sensitivity and predictive value positive
- I. Mechanics of operation

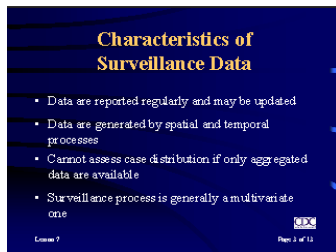
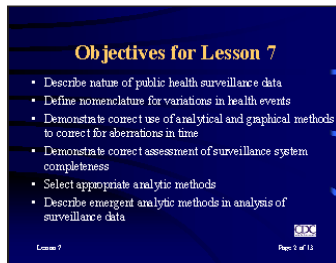
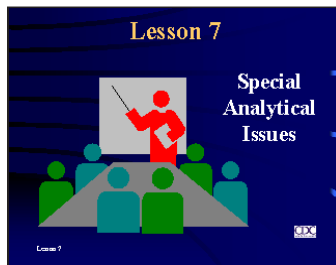
VII. Emergent methods

- A. New developments in technology and analytic methods
- B. Utility of approaches

Lesson 7

Special Analytic Issues

Content Outline



Lesson Objectives:

- **Describe nature of public health surveillance data**
 - **Define nomenclature for variations in health events**
 - **Demonstrate correct use of analytical and graphical methods to correct for aberrations in time**
 - **Demonstrate correct assessment of surveillance system completeness**
 - **Select appropriate analytical methods**
 - **Describe emergent analytical methods in analysis of surveillance data**
-

I. Nature of public health surveillance data

A. Characteristics of surveillance data

1. data are reported regularly and may be updated
 - a. methods developed for early detection of aberrations in the data should be applied as soon as provisional data are available
 - b. if analyses are implemented as part of a routine surveillance program, results can be monitored as data are updated

Return to [Lesson 7](#)

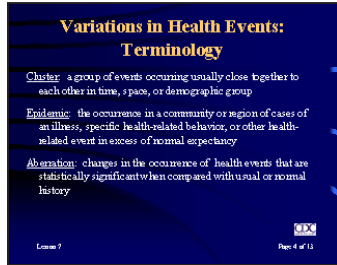
2. surveillance data are generated by a spatial as well as a temporal process
 - a. at a given point in time, cases of a disease for a given area may not appear excessive
 - b. when compared with other times or other areas at a given time, an excess may become apparent
3. when only aggregated data are available, the distribution of cases in the underlying population cannot be assessed directly
 - a. areas of aggregation are usually arbitrarily or politically defined
 - b. case definitions are not always consistent within areas
 - c. statistical inferences concerning the properties of individuals are confounded by the properties of the aggregated system
4. surveillance process is generally a multivariate one
 - a. multiple health events under surveillance may be related for a given point in time for the same area
 - b. relationship may be delayed in time for the same or nearby areas if diagnosis is uncertain or confirmation is delayed
 - c. multivariate nature of this process should be used to improve the ability of any method to detect aberrations from a baseline

II. Clustering of health events

A. Variations in health events

1. foundation of the science of epidemiology is the study of the departure of the observed patterns of the occurrence of disease from the expected pattern of occurrence

2. variations in the usual incidence of health events in different geographic areas or different time periods may provide important clues to specific risk factors or even to the etiology of the problem
3. patterns of occurrence within human populations may lead to hypotheses about the determinants of the health problem



B. Nomenclature for variations

1. cluster
 - a. a group of events occurring unusually close together
 - 1) in time and/or space
 - 2) within the same demographic group
 - b. describes uncommon events such as leukemia and suicide
 - c. tends to evoke emotional response from members of the public and / or from the media
2. epidemic
 - a. describes the occurrence in a community or region of cases of an illness, specific health-related behavior, or other health-related event clearly in excess of the expected frequency
 - b. epidemicity is relative to the usual frequency of the disease in the same area, among the specified population, at the same season of the year
 - c. influenced by agent, exposed population and its previous contact with the agent, time, and place of occurrence
 - d. the term “epidemic” evokes connotations of cholera and smallpox in the public
 - e. the term “outbreak” has less evocative connotations

- f. may require epidemiologic conditions beyond the statistical ones, such as laboratory isolates
- g. definition of epidemic may require the existence of an aberration

3. aberration

- a. describes statistical departures from a usual distribution
- b. denotes changes in the occurrence of health events that are statistically significant when compared with usual or normal history
- c. departures do not necessarily signal the “onset of an epidemic” or the “presence of a cluster”

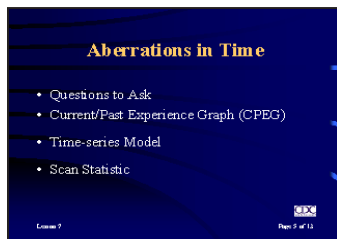
III. Aberrations in time

A. Questions to ask:

1. When does the value of reported events signal a change in the progress from past patterns?
2. How do you define past patterns?
3. Do past epidemics affect definitions of a change?
4. What other factors can cause a change?

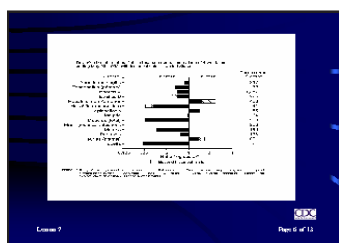
B. Graph of current and past experience

1. State health departments report the occurrence of approximately 50 notifiable diseases weekly to CDC's National Notifiable Disease Surveillance System (NNDSS)
2. List determined collaboratively by the Council of State and Territorial Epidemiologists



3. Data reported weekly in MMWR

4. CDC MMWR Current/Past Experience Graph (CPEG)

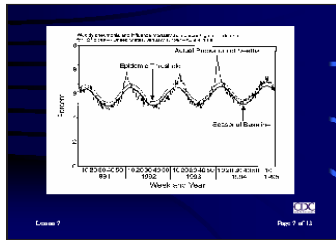


- a. compares the number of reported cases in current 4-week period for a given health event with historical data on the same condition from the preceding 5 years
- b. choice of 4 weeks as the “current period” was based on evidence that weekly fluctuation in data from disease reports usually reflects irregular reporting practices rather than actual incidence of disease
- c. use of 5 years of history achieves the objective of using the same model for all conditions portrayed
- d. modeling of data from influenza mortality surveillance has shown that more accurate forecasts are based on more recent data
- e. to increase historical sample size and account for seasonal effect, the baseline is the average for the 4 weeks preceding, the corresponding 4 week period, and the following 4 week period for the previous 5 years

5. objectives of analytic and graphical method

- a. to portray in a single comprehensible figure the weekly reports of data for approximately 20 diseases and to compare those data with past reports
- b. to highlight for further analysis the results most likely to reflect either long-term trends or epidemics
- c. objectives were formulated to reflect most recent behavior in as short a time period as possible for weekly publication, but a period long enough to assure stable results

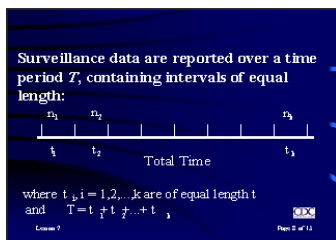
6. in-class exercise (reading or plotting a graph; optional)



C. Time-series methods

1. time-series model generally comprises components estimating the effects of secular trend, cycles, or year-to-year seasonal patterns
 - a. process of model fitting consists of identification, estimation, and diagnostic validation
 - b. then evaluate competing models on the basis of the fit of the models to the observed data and of the accuracy of the forecasts
2. Auto Regressive Integrated Moving Average (ARIMA) models
 - a. appropriate for relatively long series of data that exhibit certain regular properties over the entire series
 - b. includes terms that predict data at one point in time as a function of previous data
 - c. creates a series of averages of adjacent observations and is used to model cycles in the data
3. advantage of time-series models is that the estimation process accounts for period-to-period correlations and seasonality, as well as long-term secular trends

D. Scan statistic



1. a simple alternative to time-series method to determine significance of data
2. the maximum number of reported cases (i.e., events) in an interval of predetermined length over the time frame of interest

3. is used to test null hypothesis of uniformity of reporting against an alternative of temporal clustering
4. intended to detect relatively infrequent elevations in a series of relatively small number of events
5. in-class demonstration/exercise (calculation using scan statistic; see text page 141; optional)

IV. Aberrations in space and time

A. One method

1. assumes equal population density
2. create time-space cells
3. calculate cases of health event for each cell
4. sum maximum count for each cell to obtain test statistic

B. Knox's method - an alternative method

1. all possible pairs of cases are examined
2. each pair is classified according to whether the case-patients in the pair lived close together and had onset of the health problem (or report) close in time

C. Criticisms of Knox's method

1. choice of the critical time and space distances is arbitrary
2. makes no allowance for edge effects that arise either from natural geographic boundaries or because there are unrecorded cases outside the designated study region

D. In-class demonstration/exercise

(see pages 142,143 in text book; optional)

1. stress that there is no single test to use in all situations
2. methods such as Knox's method augment other epidemiologic methods in a systematic, integrated approach, coupled with flexibility in methods of analysis and interpretation of significance levels

V. Completeness of coverage

A. Approaches to assessment of completeness

1. one approach was developed for the enumeration of wildlife populations
 - a. used by the U.S. Census Bureau
 - b. requires two parallel surveillance systems, or a surveillance system and a survey, measuring the incidence of a single health event
 - c. provides an estimate of true total number of cases of that health event and the completeness of coverage of the two systems
2. Chandra Sekar-Deming (CSD) and Lincoln-Peterson Capture-Recapture (LPCR) methods
 - a. related approaches
 - b. benefits
 - 1) surveillance is done for a closed population
 - 2) the matching procedure successfully identifies all true matches and, conversely, only true matches are identified

Approaches to Assessment of Completeness

- U.S. Census Bureau's method
- Chandra Sekar-Deming method
- Lincoln-Peterson Capture-Recapture method

Lesson 7 Page 9 of 13

Chandra Sekar-Deming Capture - Recapture Method

| | Surveillance System 1 | | |
|-----------------------|-----------------------|--------------------|-----------|
| Surveillance System 2 | Cases Reported | Cases Not Reported | All Cases |
| Cases reported | C | N_2 | S |
| Cases not reported | N_1 | X^c | |
| All cases | R | | N |

$$N = [(R + 1)(S + 1)/(C + 1)] - 1$$

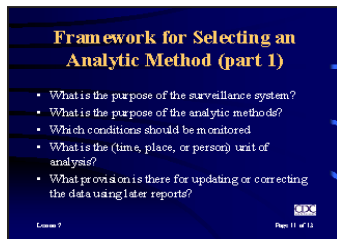
Lesson 7 Page 10 of 13

- 3) all events identified in either of the two systems are true events
- 4) the two systems are independent

3. variations on these methods can be used for public health surveillance

B. In-class demonstration/exercise

(see pages 143, 144 in textbook; optional)



VI. Selection of analytic methods

A. No single method can be used to detect all epidemics

B. Several methods can be used to help choose an analytic method

C. Purpose of surveillance system

1. patterns may be apparent from the aggregated national picture that may facilitate prevention and intervention efforts
2. data are maintained historically for the purposes of measuring trends and assessing the effects of interventions

D. Purpose of analytic method

1. a single method cannot be expected to distinguish between a change in historical trend and a onetime epidemic with unsustained increases
2. analyst must identify purpose of analysis before choosing an analytic method
3. if nature of data is determined and the questions are well-defined, the results of the analytic method can be used to augment other sources of information

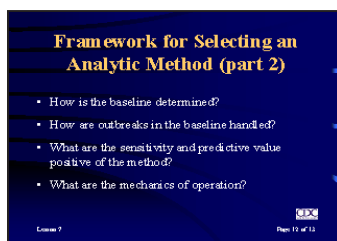
[Return to Lesson 7](#)

4. purpose of Current/Past Experience Graph (CPEG)
 - a. to facilitate routine analysis of surveillance data and to supplement other sources of information
 - b. most useful for conditions without long-term historical trends

E. Selection of conditions for monitoring

(using CPEG methodology to illustrate concepts)

1. routine analyses should be used for conditions for which there are public health interventions
2. CPEG methodology is used for national surveillance of notifiable conditions to facilitate prevention and intervention efforts
 - a. is useful for conditions that occur often enough so that one or two cases do not constitute a significant flag
 - b. limitations
 - 1) if the surveillance data are not already analyzed for trend and period effects
 - 2) if the variance of the numerator (present cases) cannot be assumed to have the same variance as the observations in the denominator (baseline data)
 - 3) if the series exhibits considerable correlation for first-order (adjacent) observations
 - 4) for rare conditions, the instability caused by small numbers of reported cases may make the results unsuitable for repeated use



F. Units of analyses

1. for CPEG, national data were selected
2. short and recent time period for weekly publication
3. makes results useful for timely intervention

G. Provision for updating/correcting data

1. cases are reported as early as possible and then later confirmed or modified
2. methodology of CPEG is applied to provisional (earliest) data

H. Determination of baseline

1. choice of 5 years as baseline was based on a consideration of appropriate sample size balanced by a desire to use the same method for all conditions
2. epidemics or changes in trend in baseline will increase the variance of the baseline and thus offset any benefit of additional data
3. additional source of variation may be increased in reporting due to intensive investigation

G. Dealing with epidemics in baseline

1. CPEG does not adjust for epidemics in the baseline
2. result is a progressive decline in sensitivity
3. to address this, one could use a median of the baseline reports (rather than a mean)
 - a. replacement may require altering the technique used to compute the point for signaling aberrations

- b. alternative methods for calculating this are not as accessible to the practicing epidemiologist as the CPEG methodology

J. Sensitivity and predictive value positive

1. application of CPEG by states showed a sensitivity of 74%
2. application of CPEG by states showed positive predictive value of 52%

K. Mechanics of operation

1. any method must be easily implemented in the routine work of the practicing epidemiologist
2. software is being developed so that health departments can generate CPEG locally

VII. Emergent methods

A. New developments in technology and analytic methods

1. group of Bayesian methods
 - a. offers mechanism to include in the analysis information beyond the data being collected
 - b. empirical Bayes procedure for increasing stability of observed rates from areas with small populations



B. Utility of approaches

1. depends on collaboration among scientific disciplines
- 2 . statistical development must consider the unique aspects of surveillance data
3. epidemiologic practice must incorporate the sound basis of statistical science